

KEKER
VAN NEST
& PETERS

LATHAM & WATKINS LLP

MORRISON FOERSTER

January 16, 2025

Hon. Ona T. Wang
 Daniel Patrick Moynihan
 United States Courthouse
 500 Pearl St.
 New York, NY 10007-1312

cc: All Counsel of Record Line (via ECF)

Re: *Authors Guild et al. v. OpenAI, Inc., et al.*, Case Nos. 1:23-cv-08292-SHS & 1:23-cv-10211-SHS [ECF 306, 262 – Plaintiffs’ Letter Brief re Source Code]

Dear Magistrate Judge Wang:

Plaintiffs’ motion to compel OpenAI to make available for inspection nine overbroad categories of “source code and related information”—weeks before responses to corresponding RFPs requesting the code are even due—is an improper attempt to circumvent the rules of discovery and the Court’s prior rulings. Worse, it wastes the Court’s time.

As OpenAI has informed Plaintiffs, it is already in the process of collecting and preparing relevant source code for inspection. This includes the code used in ChatGPT and the training of the underlying GPT models the Court has determined to be at issue. *See* Dkt. 293 at 2. However, until the filing of its letter brief, Plaintiffs had not issued any discovery request specifically targeting source code. In fact, Plaintiffs only served the RFPs requesting the source code at issue in this motion on January 10—**six days ago**. *See* Ex. 1. OpenAI will respond to these January 10 RFPs in due course and, consistent with its prior representations, intends to make available for inspection relevant source code for ChatGPT and the training of the GPT models at issue. OpenAI **does not object to producing relevant source code**. It does, however, object to Plaintiffs’ improper demand for a premature commitment on the scope of production before OpenAI has had any opportunity to assert objections or engage in a meaningful meet-and-confer process.

Plaintiffs, perhaps recognizing that their January 10 RFPs are not ripe, attempt to link their new demands to a hodge podge of earlier RFPs that do not seek source code at all. Plaintiffs only identified these earlier RFPs on January 9—two business days before filing this motion—effectively denying OpenAI an opportunity to evaluate or respond to their contentions before this motion was filed. That is not how the meet-and-confer process is supposed to work.

For the foregoing reasons, there is no ripe dispute over source code for the Court to resolve. Accordingly, the Court should deny Plaintiffs request as premature.

I. Plaintiffs have ignored their obligation to meet-and-confer regarding source code.

Plaintiffs have served more than 130 requests for production. Until a week ago, *Plaintiffs did not contend* that any of its RFPs encompassed source code in meet-and-confer discussions. Despite

this, OpenAI has been willing to produce source code. In particular, OpenAI voluntarily agreed—subject to its objections—to produce certain source code in response to an RFP asking for “Documents sufficient to show the number of reproductions made of training data in the course of training [OpenAI’s] LLMs.” Specifically, OpenAI agreed that “for text training datasets for models used for ChatGPT, [it] will make available for inspection the source code related to processing that data for use during pre-training or post-training and filtering that data, to the extent such code can be located pursuant to a reasonable search.” Dkt. 306-3 at pp. 13-14. On November 18, OpenAI made available a substantial portion of this agreed-upon code, and Plaintiffs reviewed that code on December 18.

On December 26—while OpenAI’s counsel and employees were out for the holiday break—Plaintiffs’ counsel sent an email alleging “deficiencies” in OpenAI’s code production and, for the first time, asserted 11 new categories of source code that they now seek to compel. Ex. 2. In response, OpenAI expressed willingness to meet and confer over the extent to which these new source code categories were responsive to Plaintiffs’ existing RFPs. However, Plaintiffs failed to identify *any* RFPs supporting their 11 new categories at that time. *Id.* The parties met for a meet-and-confer call on January 9, and Plaintiffs’ counsel then orally identified (for the first time) 14 RFPs—including 5 not mentioned in Plaintiffs’ brief—that they contended cover the newly identified categories of source code. OpenAI agreed to consider Plaintiffs position.

The next day, on January 10, Plaintiffs served its Ninth Set of Requests for Production, which finally requested the source code now at issue. OpenAI reasonably understood this to have resolved the issue, as the new RFPs provided a framework for properly addressing Plaintiffs’ requests under the Federal Rules. Exs. 1 & 2. Surprisingly, Plaintiffs filed this motion the following business day.

II. Plaintiffs’ request is premature and circumvents the FRCP and the Court’s rulings.

OpenAI’s responses to Plaintiffs’ Ninth Set of RFPs are not due until February 10. Until OpenAI responds, and the parties have had an opportunity to meet-and-confer, there is no dispute for the Court to resolve. Plaintiffs’ attempt to circumvent the rules of discovery is improper, and the request should be denied. *Stokes v. Wayne Cnty.*, 2023 WL 1775637, at *2 (W.D.N.Y. 2023) (holding that motion to compel filed before response due “must be denied as premature”).

Moreover, Plaintiffs’ new RFPs are objectionable for numerous reasons, including their overbreadth and improper scope. For example, the RFPs ostensibly cover source code and other purportedly “related information” used in research, experiments, and other models that the Court has already ruled outside the scope of discovery. *See* Dkt. 293 at 2. To take just one example, its purported source code category “H” **does not seek source code at all**. Instead, it surreptitiously requests all “training datasets present and referenced in the source code that have not been produced.” Dkt. 306 at 2. But OpenAI has already agreed to make available for inspection relevant training datasets actually used to train the models at issue in this case. So Plaintiffs’ new request seeks training datasets that the Court has already found to be outside the

scope of discovery. Plaintiffs should not be permitted to re-litigate this issue—and certainly not through an afterthought request that Plaintiffs do not even substantively address in their brief.¹

III. Plaintiffs' cited RFPs (and OpenAI's responses to them) do not cover source code.

Although OpenAI has never had the opportunity to meet-and-confer about whether the nine RFPs cited in Plaintiffs' brief cover source code, the text of the requests make clear that they do not. For this reason, OpenAI did not agree to produce source code in response to these requests when it served its operative responses to each of these requests at least five months ago. Exs. 3-5; Dkt. 306-3. Those responses were the product of *extensive* meet-and-confer efforts based on Plaintiffs prior proffers of relevance. Yet Plaintiffs never once during these many meet-and-confer efforts quarreled with OpenAI's responses as improperly excluding source code.

This is unsurprising, as the specific RFPs identified by Plaintiffs are not directed at source code with any reasonable particularity. *See Royal Park Invs. SA/NV v. Deutsche Bank Nat'l Tr. Co.*, No. 14CV04394AJNBCM, 2016 WL 4613390, at *10 n.8 (S.D.N.Y. Aug. 31, 2016) (observing that requests broadly worded requests asking for all documents “concerning” or “related to” an issue “plainly violate Fed. R. Civ. P. 34(b)(1)(A), which requires a party to ‘describe with reasonable particularity’ each item or category to be produced.”). For example, RFP Nos. 17, 18, 31, 39, 55, and 65 are all directed to categories of documents sufficient to identify OpenAI's knowledge of the specific text training data, or types of text training data, used to train OpenAI's ChatGPT models; source code is simply not responsive to these requests as written. RFP No. 41 is directed at internal documents discussing or describing the methods and techniques it uses to prevent regurgitation of copyrighted content; RFP 60 is directed to documents explaining how OpenAI itself describes the training process (e.g., through use of metaphor), and RFP 65 is directed towards OpenAI's approach to robots.txt. Exs. 3-5. Again, none of these requests ask OpenAI to produce source code.

In sum, none of Plaintiffs' existing RFPs are directed at source code. Plaintiffs' first, and only, discovery requests seeking source code are the RFPs propounded on January 10 that mirror the categories of source code identified in the motion. OpenAI agrees that it is producing source code in this case, but the contours of what source code it agrees to produce should be the product of the discovery process—*i.e.*, OpenAI's responses to Plaintiffs' January 10 RFPs and any subsequent meet-and-confer (if necessary).

* * * * *

For the foregoing reasons, the Court should deny Plaintiffs' requests for source code and other related information—at the very least, as premature.

¹ Relatedly, Plaintiffs' category “I” vaguely demands “documents referenced in source code that relate to the source and copyrightability of training data.” As noted in the meet-and-confer correspondence, Ex. 2, OpenAI does not know what documents Plaintiffs are referencing.

Respectfully,

KEKER, VAN NEST & PETERS
LLP

LATHAM & WATKINS LLP

MORRISON & FOERSTER LLP

/s/ Edward A. Bayley

/s/ Elana Nightingale Dawson

/s/ Andrew L. Perito